

On the Use of Textural Features for Writer Identification in Old Handwritten Music Scores

Alicia Fornés, Josep Lladós, Gemma Sánchez
Computer Vision Center - Dept. of Computer Science
Universitat Autònoma de Barcelona
Edifici O, 08193, Bellaterra, Spain
{afornes,josep,gemma}@cvc.uab.cat

Horst Bunke
Institute of Computer Science and Applied Mathematics
University of Bern, Neubrückstrasse 10
CH-3012 Bern, Switzerland
bunke@iam.unibe.ch

Abstract

Writer identification consists in determining the writer of a piece of handwriting from a set of writers. In this paper we present a system for writer identification in old handwritten music scores which uses only music notation to determine the author. The steps of the proposed system are the following. First of all, the music sheet is preprocessed for obtaining a music score without the staff lines. Afterwards, four different methods for generating texture images from music symbols are applied. Every approach uses a different spatial variation when combining the music symbols to generate the textures. Finally, Gabor filters and Grey-scale Co-occurrence matrices are used to obtain the features. The classification is performed using a k -NN classifier based on Euclidean distance. The proposed method has been tested on a database of old music scores from the 17th to 19th centuries, achieving encouraging identification rates.

1 Introduction

In the field of document analysis, there is an increasing interest in historical documents, towards the inclusion of these documents into digital libraries, helping in the diffusion and preservation of artistic and cultural heritage. Optical Music Recognition (OMR) is a classical area of interest of Document Image Analysis and Recognition (DIAR) that combines textual and graphical information. In addition to the preservation in digital format, the interest of applying DIAR to historical handwritten musical scores is twofold:

the transcription of the score to a standard format, and the classification of the document in terms of the writer. In fact, many historical archives contain a huge number of sheets of musical compositions without information about the composer, and musicologists have difficulties for identifying the writer (or the copier) of every sheet. For that reason, a system for writer identification in old music scores could help musicologists in such a time consuming task.

Writer identification in handwritten text documents is a mature area of study (see [9], [10], [11]), whereas the identification of the writer of music scores is still an open problem. To the best of our knowledge, only one project (see [1], [4]) has been performed about writer identification in music scores. The authors have developed a prototype that analyzes the music score and then extracts some features about structural information of the music symbols and notes. However no quantitative results have been published, and as far as we know, this work has not been continued.

Most compositions in previous centuries were sacred music, containing lyrics (text) for singers. In these scores, the writer identification methods for handwritten text documents could be applied for lyrics. However, the aim of our work is to evaluate the performance of writer identification methods extracting features only from music symbols, because of the following reasons. Firstly, graphic notation can be more distinctive (more personal) than text, and secondly, in some cases the writer of the music symbols and the writer of the lyrics are not the same person. Moreover, our methodology will also be useful for writer identification in generic music scores, including music scores for instruments, without any text.

In this paper we present an off-line text-independent approach for performing writer identification in musical scores, which avoids the recognition of the elements in the score. Consequently the system will be faster and more robust, avoiding the dependence on a good music recognizer.

In [3] we presented an approach for writer identification using 100 features extracted from music lines. Those features were derived from connected components, enclosed regions, contours, fractals and basic measurements. The experimental results using those local features were quite good, but in some cases a single music line has not enough information to correctly identify the writer. In order to achieve a more reliable writer identification, we propose the use of textural features, because they are able to represent the music score globally rather than focusing on a set of predefined local features. In [7], texture images are generated from printed text for script and language identification. In [9], texture images are generated from handwritten text for writer identification. Both approaches demonstrate that textural features can be successfully used for writer and script identification. In the current paper we have adapted part of the writer identification approach described in [9] to old musical scores, where instead of letters of the alphabet, music notations are used for generating textures, and consequently, textural features can be obtained for the identification of the writer.

The remainder of this paper is structured as follows. In the next section the preprocessing steps are presented, and in Section 3 the feature extraction approach is described. Experimental results are presented in Section 4. Finally, Section 5 concludes the paper and proposes future work.

2 Preprocessing

The preprocessing phase consists in binarizing the image, removing staff lines and generating the texture images from music notes. The process is described next.

2.1 Binarization and Staff removal

The input gray-level scanned image (at a resolution of 300 dpi) is first binarized with the adaptive binarization technique proposed by Niblack [6]. Then, filtering and morphological operations are applied to reduce noise. Afterwards, the image is deskewed using the Hough Transform. Then, the staff lines are removed, because they are usually printed, and consequently, they are not useful for writer identification. The staff removal process is difficult because of paper degradation, distortions, gaps and the warping effect. The method consists in obtaining a coarse approximation of the staff lines applying median filters with a horizontal mask and then reconstructing the staff joining these segments. Afterwards, a contour tracking process is used

for following and removing every staff line, taking into account the coarse approximation when gaps are appearing. For further details, see [2].

2.2 Generation of music textures

Once the music symbols have been segmented, the music symbols are used for generating texture images. It must be said that textural features directly computed on the music score without any staff removal are not effective, because the frequency of the staff lines affects to the values of the textural features. We have applied four different methods for obtaining the texture images (see Fig. 1). Each method is characterized by a different spatial variation when combining the music symbols to generate the textures. In all cases, the size of the texture image is of 2048x2048 pixels. The following four different methods for obtaining the textures have been applied:

- **Basic Texture:** consists in taking all the music symbols obtained in the previous step, without any other processing (see Fig. 1(a),(b)). In this way, the music symbols appear in the same order than in the music score, keeping the inter-symbol distance.
- **TextLine Texture:** consists in taking randomly music symbols and putting them in a reference line, with the same inter-symbol distance (see Fig. 1(c),(d)). In this way, if the music score contains a group of the same kind of music symbol (i.e. quarters or rests), they will be randomly distributed over the texture, achieving texture independence of the rhythm.
- **AspectRatio Texture:** consists in taking the idea of TextLine texture, but making all the symbols of equal size (see Fig. 1(e),(f)). For every symbol that must be resized, its aspect ratio will be maintained. The main purpose is to avoid gaps in the texture, obtaining a higher density of the texture.
- **Resize Texture:** consists in the same idea as AspectRatio Texture, but without the preservation of the aspect ratio in the resizing process (see Fig. 1(g),(h)). In this way, the appearance of the symbol is distorted (symbols are taller comparing to the original shape).

One can see that the first two approaches generate texture images that look like a music score, whereas the last two approaches generate more compact and synthetic texture images. In fact, the AspectRatio texture has the inconvenient of creating some big black areas, because small compact music symbols (such as dots and half rests) are extremely enlarged. It is important to remark that in all texture images, the first writer can be easily distinguished from the second one, because the first writer tends to use more curves than straight lines (see Figure 1).

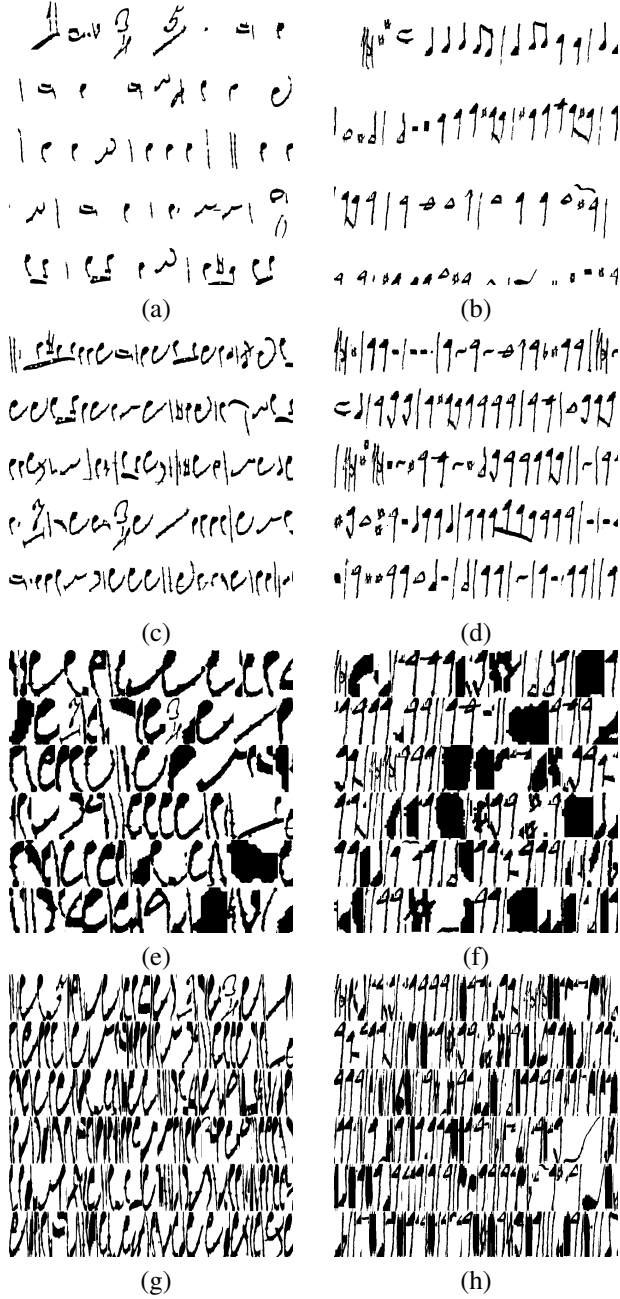


Figure 1. Texture images generated from music lines. The first column shows textures from writer 1, the second column corresponds to textures from writer 2: (a),(b) Basic textures; (c),(d) TextLine textures; (e),(f) AspectRatio textures, (g),(h) Resize textures.

3 Feature extraction from textures

Textural features can be used for analysing the global frequency content of the data. In [9] and [7], texture images

are generated from text, in order to compute textural features. We have been inspired by this idea, generating music texture images for being able to extract the following textural features: Gabor features and Gray-Scale co-occurrence matrices.

3.1 Gabor features

The multi-channel Gabor filtering technique [12] can be seen as a window Fourier Transform in which the window function is Gaussian. This technique is based on the psychophysical findings that affirm that the processing of pictorial information in the human visual cortex involves a set of parallel and quasi-independent cortical channels. Every cortical channel can be modeled by a pair of Gabor filters $h_e(x, y; f, \theta)$ and $h_o(x, y; f, \theta)$. These filters are of opposite symmetry and are computed as:

$$\begin{cases} h_e(x, y; f, \theta) = g(x, y) \cos(2\pi f(x \cos \theta + y \sin \theta)) \\ h_o(x, y; f, \theta) = g(x, y) \sin(2\pi f(x \cos \theta + y \sin \theta)) \end{cases} \quad (1)$$

where $g(x, y)$ is a 2D Gaussian function, the central frequency is f , and θ is the orientation which define the location of the channel in the frequency plane. Afterwards, the Fourier transform (FFT) of the filters are computed as:

$$\begin{cases} q_e(x, y) = FFT^{-1} [P(u, v) H_e(u, v)] \\ q_o(x, y) = FFT^{-1} [P(u, v) H_o(u, v)] \end{cases} \quad (2)$$

where $P(u, v)$ is the Fourier Transform of the input image $p(x, y)$ and $H_e(u, v)$ and $H_o(u, v)$ are the Fourier Transform of the filters $h_e(x, y; f, \theta)$ and $h_o(x, y; f, \theta)$; respectively. Finally, we perform a combination of the two filters, and a single value at each pixel is obtained:

$$q(x, y) = \sqrt{q_e^2(x, y) + q_o^2(x, y)} \quad (3)$$

For the computation of features, we have to define the angle θ and the central frequency f , which specify the location of the Gabor filter on the frequency plane. In [12], it has been shown that for an image of size $N \times N$, the important frequency components are found within $f \leq N/4$ cycles/degree. For this reason, the parameters used are $f = \{4, 8, 16, 32\}$ and $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. The output corresponds to $4 \times 4 = 16$ images. Extracting the mean and the standard deviation we obtain $16 \times 2 = 32$ features.

3.2 GSCM features

The Grey-Scale Co-occurrence Matrices (GSCM) are typically used for describing grey scale textures (see [7]). If an image contains N grey levels, for every distance d and angle θ we obtain a matrix $N \times N$ defined as $GSCM_{d, \theta}$, where $GSCM_{d, \theta}(a, b)$ corresponds to the number of pairs

($P1, P2$) where $P1$ is of grey value a , $P2$ is of grey value b , and $P1$ and $P2$ are separated by distance d and angle θ .

Whereas GSCM are of a high computational cost for grey level images, they are fast to compute for binary images, because there are only two grey values. The parameters used in our method are the distance $d = \{1, 2, 3, 4, 5\}$; and the orientation $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. The output corresponds to 20 matrices of dimension 2×2 , and due to the diagonal symmetry, there are only 3 independent values in each matrix. In total we obtain $20 \times 3 = 60$ features.

4 Experimental Results

We have tested our method with 200 music pages¹ from 20 different writers, where every writer has written 10 pages. Due to the large amount of symbols on every music page, three different texture images can be generated per page, obtaining a database of $20 \times 10 \times 3 = 600$ music textures. These pages are extracted from a collection of music scores of the 17th, 18th and 19th centuries, which have been obtained from two archives in Catalonia (Spain): the archive of Seminar of Barcelona and the archive of Canet de Mar.

For the experiments, we have used 5 test subsets, randomly chosen, containing one page per writer. This means that all the three music textures obtained from every page are used in the test set. Due to the importance of the obtention of independent test subsets, all the three textures generated from one music page belong to the same subset. For each test subset or 60 images, the remaining 540 images are used for training. The classification has been performed using a 5-Nearest Neighbor classifier based on Euclidean distance and cross validation. The chosen value $k=5$ has been empirically determined to be the optimal choice. Due to the fact that every music page generates three texture images, the three texture images must be only assigned to one class. This is performed combining the classification results of the three texture images using the Borda count method.

In Table 1 the writer identification rates (w.i.r.) for the Basic, Textline, AspectRatio and Resize textures are shown. One can see that Resize textures obtain the highest w.i.r. (73%) using the combination of Gabor and GSCM features, whereas Basic and AspectRatio reach lower identification rates, even when the features are not combined (GSCM features reach a w.i.r. of 58% in Basic textures and Gabor features reach a w.i.r. of 65% in AspRatio textures). Notice that the textures extracted using the TextLine method obtain in all cases the lowest rates.

We have compared the proposed method with a previous work based on the extraction of 100 typical features for handwritten text recognition (see [3]), and we can affirm that textural features reach higher identification rates. In Table

Table 1. Writer identification rates using Gabor and GSCM features for the four methods applied for obtaining texture images.

Features	Basic	A.Ratio	TextLine	Resize
Gabor	53%	65%	34%	64%
GSCM	58%	55%	45%	66%
Gabor & GSCM	55%	52%	47%	73%

2, a comparison of the results using both methods is shown. A database of music pages from 7 writers reach a w.i.r. of 79% when classifying using 100 text line features extracted from music lines, whereas a database of music pages from 10 writers (three writers more than in the previous one) the method which uses textural features applied to Resize texture images reach a w.i.r. of 84%. In addition, it is important to notice that the writer identification rate decreases significantly when adding more writers to the database (from 92% with 5 writers to 73% with 20 writers) because the different writer styles become very close. In fact, the confusion matrices analyzed show that the disciples of the same musician (or that belong to the same place and time period) tend to have a very similar writer style (see Figure 2).

Table 2. Classification Results: Writer identification rates using 100 line features and textural features for different database sizes.

Experiment	N. of Writers	W.I.Rate
100 line features	7	79%
92 textural features	5	92%
92 textural features	10	84%
92 textural features	15	73.3%
92 textural features	20	73%

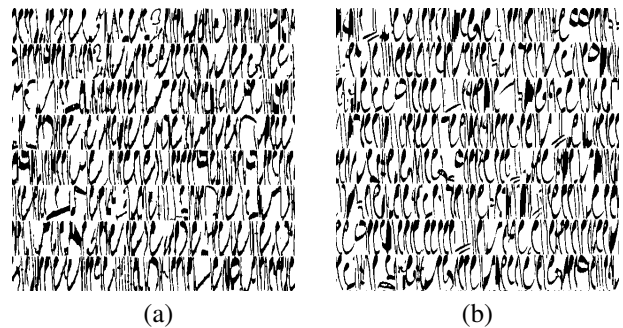


Figure 2. Resize texture images from two writers: Both texture images are very similar although they belong to different classes.

¹This dataset is available under request to the authors of this paper.

The suitability of the textural features has also been analyzed, because some of them could be unnecessary or even redundant. The goal of feature selection is to find the best subset of features that perform better than the original ones. We have performed the Sequential Forward Search (SFS), Sequential Backward Search (SBS), Sequential Floating Forward Search (SFFS), Sequential Floating Backward Search (SFBS) (see [5], [8]). For the experiments, one of the five subsets is used as the test set and the others as the prototypes in the 5-NN classifier. To evaluate the quality of a selected feature subset, iteratively three subsets are used in the classifier and the remaining set is used to measure the quality of the feature subset under consideration. Once the algorithm finds the best feature subset, the fifth subset is used for the final writer identification rate.

Table 3. Writer identification rates using Feature Set Search methods.

Experiment	N. of Features	W.I.Rate
Gabor & GSCM	92	73%
SFS	32	70%
SBS	18	65%
SFFS	28	70%
SFBS	11	70%

In Table 3 results of Resize textures of feature selection algorithms are shown. The first row again shows the baseline rate, and the next ones show the results using SFS, SBS, SFFS and SFBS feature set search methods. It is important to remark that they do not improve the writer identification rate. This fact shows that there are not many dependent or irrelevant features in the original feature set, being all the features important for the classification. Notice that these selected features are specific to this database, and the results could potentially be quite different for other datasets.

5. Conclusions

In this paper we have presented a new method for writer identification in musical scores which computes GSCM and Gabor features to texture images generated from music symbols. Although Resize textures are the ones with the highest classification rates (even when the writer styles are very similar), the work can be extended if the textural features obtained from the four different approaches are combined in a single vector, so that the feature selection methods could possibly increase the final classification rate.

We are currently working on the combination of the information extracted from texture images and the information extracted from music lines (described in [3]), and also the addition of specific features for musical notation to the current set of features.

Acknowledgements

We would like to thank Prof. J.M.Gregori for his help in accessing the historical archives, and V.Kilchherr and Dr. A.Schlapbach for providing support in the experiments. This work has been partially supported by the Spanish projects TIN2006-15694-C02-02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018) and 2007 BE-1 00086.

References

- [1] I. Bruder, T. Ignatova, and L. Milewski. Integrating knowledge components for writer identification in a digital archive of historical music scores. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)*, pages 397–397, New York, NY, USA, 2004. ACM.
- [2] A. Fornés, J. Lladós, and G. Sánchez. Primitive segmentation in old handwritten music scores. In W.Liu and J.Lladós, editors, *Graphics Recognition: Ten Years Review and Future Perspectives*, volume 3926 of *Lecture Notes in Computer Science*, pages 279–290. Springer, 2006.
- [3] A. Fornés, J. Lladós, G. Sánchez, and H. Bunke. Writer identification in old handwritten music scores. In *8th International Workshop on Document Analysis Systems*, pages 347–353, Nara, Japan, September 2008.
- [4] R. Göcke. Building a system for writer identification on handwritten music scores. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA)*, pages 250–255, Rhodes, Greece, 30 June – 2 July 2003.
- [5] J. Kittler. Feature set search algorithms. In C.H.Chen, editor, *Pattern Recognition and Signal Processing*, 1978.
- [6] W. Niblack. *An Introduction to Digital Image Processing*. Prentice Hall, 1986.
- [7] G. Peake and T. Tan. Script and language identification from document images. *Document Image Analysis, 1997. (DIA '97) Proceedings., Workshop on*, pages 10–17, Jun 1997.
- [8] P.Pudil, J.Novovicová, and J.Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
- [9] H. Said, T. Tan, and K. Baker. Personal identification based on handwriting. *Pattern Recognition*, 33(1):149–160, January 2000.
- [10] A. Schlapbach and H. Bunke. Off-line writer identification and verification using gaussian mixture models. In S. Marinai and H. Fujisawa, editors, *Machine Learning in Document Analysis and Recognition*, volume 90 of *Studies in Computational Intelligence*, pages 409–428. Springer, 2008.
- [11] L. Schomaker and M. Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):787–798, June 2004.
- [12] T. Tan. Written language recognition based on texture analysis. *Image Processing, 1996. Proceedings., International Conference on*, 1:185–188 vol.2, Sep 1996.